Heliyon 10 (2024) e32934

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

An end-to-end gait recognition system for covariate conditions using custom kernel CNN

Babar Ali [a], Maryam Bukhari [a], Muazzam Maqsood [a], Jihoon Moon [b], Eenjun Hwang [c], Seungmin Rho [d,*]

[a] Department of Computer Science, COMSATS University Islamabad, Attock Campus, Pakistan
[b] Department of AI and Big Data, Soonchunhyang University, Asan, 31538, South Korea
[c] School of Electrical Engineering, Korea University, Seoul, 02841, South Korea
[d] Department of Industrial Security, Chung-Ang University, Seoul, 06974, South Korea

ARTICLE INFO

Keywords:
Gait recognition
Covariate factors
Deep learning
Convolutional neural networks
Custom kernel CNN

ABSTRACT

Gait recognition is the identification of individuals based on how they walk. It can identify an individual of interest without their intervention, making it better suited for surveillance from afar. Computer-aided silhouette-based gait analysis is frequently employed due to its efficiency and effectiveness. However, covariate conditions have a significant influence on individual recognition because they conceal essential features that are helpful in recognizing individuals from their walking style. To address such issues, we proposed a novel deep-learning framework to tackle covariate conditions in gait by proposing regions subject to covariate conditions. The features extracted from those regions will be neglected to keep the model's performance effective with custom kernels. The proposed technique sets aside static and dynamic areas of interest, where static areas contain covariates, and then features are learnt from the dynamic regions unaffected by covariates to effectively recognize individuals. The features were extracted using three customized kernels, and the results were concatenated to produce a fused feature map. Afterward, CNN learns and extracts the features from the proposed regions to recognize an individual. The suggested approach is an end-to-end system that eliminates the requirement for manual region proposal and feature extraction, which would improve gait-based identification of individuals in real-world scenarios. The experimentation is performed on publicly available dataset i.e. CASIA A, and CASIA C. The findings indicate that subjects wearing bags produced 90 % accuracy, and subjects wearing coats produced 58 % accuracy. Likewise, recognizing individuals with different walking speeds also exhibited excellent results, with an accuracy of 94 % for fast and 96 % for slow-paced walk patterns, which shows improvement compared to previous deep learning methods.

1. Introduction

Human identification, often known as recognition, is the process of uniquely identifying individuals based on their biometric traits. Various biometric features can be used for this purpose, with each offering unique information for identification. The most common

* Corresponding author.
E-mail address: smrho@cau.ac.kr (S. Rho).

https://doi.org/10.1016/j.heliyon.2024.e32934
Received 11 November 2023; Received in revised form 18 May 2024; Accepted 12 June 2024
Available online 16 June 2024

biometric features employed for recognition include fingerprint scanning, iris recognition, and face recognition [1]. However, all these biometric features require close contact with the subject, which is impossible in some scenarios. On the other hand, a person's gait does not require human intervention and can be acquired and identified from a longer distance utilizing their walking style. More precisely, gait refers to the distinctive walking pattern of an individual, encompassing various aspects of locomotion such as stride length, cadence, posture, and rhythm. It serves as a unique biometric identifier due to its inherent individuality and potential for recognition. Gait analysis has been playing a great role in recognizing gait patterns and identifying the person of interest from a greater distance [2]. Gait recognition also demonstrates potential in areas where traditional biometrics may not be applicable or feasible. When individuals are too far away or facial features cannot be recognized, gait analysis can still give useful information for identification. Gait recognition, unlike other biometrics such as fingerprints or iris scans, can be carried out without human intervention. This characteristic of gait makes it more ideal for surveillance systems [3]. Computer vision-based silhouette analysis methods have been utilized in past studies to identify gait patterns and classify individuals based on motion analysis. However, like any other biometric feature, gait recognition has also some limitations. Environmental factors, variations in walking speed, and different camera viewpoints can affect the accuracy of gait recognition systems [4].

Gait recognition offers a unique and non-intrusive biometric feature that can be used for identification and verification purposes. More precisely, gait analysis techniques are divided into two main categories: model-based and appearance-based techniques. Model-based gait recognition is an approach that utilizes mathematical models to represent the underlying structure and properties of human gait. These models explain how different body parts work together while walking [5]. By leveraging these models, these techniques aim to capture and analyze the unique characteristics of an individual's walking patterns. The models employed in model-based gait recognition can describe various aspects of gait, including joint angles, joint torques, muscle activations, and spatiotemporal relationships between body parts [6]. One of the key advantages of model-based gait recognition is its ability to provide a deeper understanding of the underlying mechanisms of human locomotion. By comparing the derived model parameters, gait recognition systems can identify individuals based on the unique characteristics of their gait models. Model-based approaches use body parameters as features, hence also needing high-resolution videos [7]. On the other hand, appearance-based gait recognition is a popular approach that focuses on extracting and analyzing visual features from gait sequences to identify individuals. These methods primarily utilize the visual appearance of a person's gait, including the shape, and motion characteristics of body parts during walking. Appearance-based techniques use directly extracted features from image templates or sequences of gait. These features are often obtained through video recordings or depth sensors [8]. Video recordings are normally divided into frames so that important temporal information is kept, and then those frames are processed further for feature extraction. They have also utilized silhouettes that contain crucial information about the shape and posture of the human body and are extracted from video frames, also known as gait sequences [3]. Unlike model-based techniques, appearance-based methods work with low-resolution videos, making them more suitable and robust for real-world scenarios.

Gait sequences in appearance-based models are represented by features from the frequency domain, silhouettes, chrono-gait images, and gait energy images (GEIs). The GEI is a widely used representation in gait recognition that captures the spatiotemporal information of an individual's gait pattern. These GEIs are proposed by Han et al. [9] in which they create a single image averaged or normalized over silhouette sequences of the gait cycle, containing important temporal information. It is less prone to segmentation errors or information loss. The GEI is obtained by summing the energy values across multiple frames of a gait sequence, resulting in a single representative image. However, challenges such as variations in clothing, viewpoint shifts, and occlusions can impact the

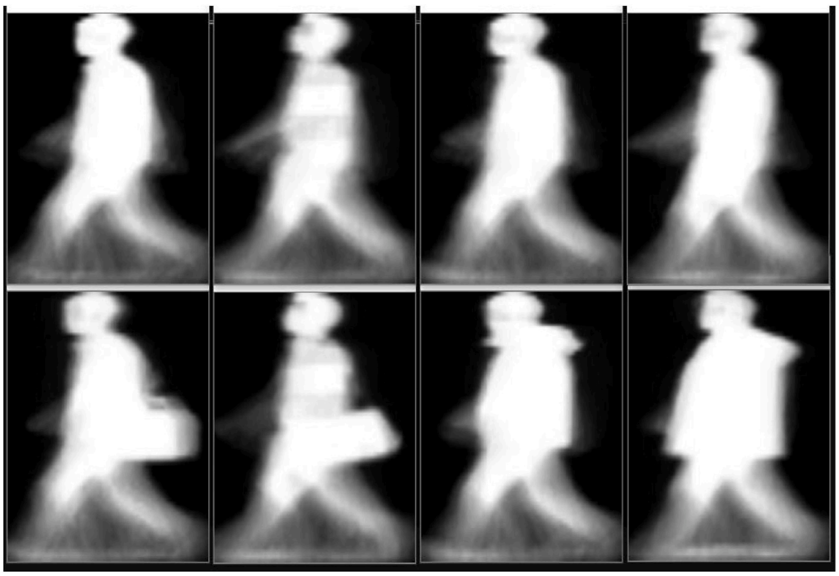

Fig. 1. The subjects in the upper row walked normally, whereas those in the bottom walked with covariate factors such as bags and coats.

accuracy of automated gait analysis [10]. Ongoing research aims to overcome these challenges and further enhance the accuracy and robustness of appearance-based gait recognition systems. According to researchers, matching GEI templates when covariates are not present shows significant results, but covariates are mostly present in real-world scenarios, making it a difficult challenge [11]. A covariate refers to an external factor or condition introduced to modify an individual's natural gait pattern and restrict their movement in a specific manner. An example of subject's GEIs with a normal walk and wearing bags and coats is depicted in Fig. 1. More specifically, when a person with a carrying condition occurs, such as wearing a long coat, pace variation, or carrying a bag, it is known as a covariate condition [12].

It is observed from Fig. 1 that, carrying covariate factors disrupts static regions or non-moving areas of the human body, like when carrying a backpack or wearing a long coat. On the other hand, the dynamic areas contain more information about the gait patterns, making them more important for gait recognition [13]. When the model is trained on a normal walk sequence (GEI) but tested on different covariate conditions, it is called an unknown or strict covariate condition. However, if the model gets trained and tested on similar covariates, such as bags or coats, those are known as known covariate conditions [14]. To address such issues, this study proposed a novel deep learning model and has the following contributions.

- A CNN-based method is proposed to efficiently handle known and unknown covariate conditions utilizing GEI.
- Introducing parallel-running custom kernels to extract dynamic and static feature maps from the input GEI images
- Experimentation on large benchmark datasets, including CASIA B and CASIA C, indicates that the proposed model performs well even with unknown covariate conditions.

The rest of the paper is organized as follows: Section 2 provides a comprehensive literature review, Section 3 introduces the proposed methodology, Section 4 discusses the experimental findings, and finally, Section 5 concludes the research by discussing future implications and potential routes for further exploration.

2. Literature review

2.1. Model-based methods

A gait recognition approach was introduced in Ref. [15], also known as Gait Graph, in which, to achieve gait recognition, the 2D human pose that was taken as input and the graph convolutional network were merged. The graph network was optimized using supervised contrast loss, whereas the 256-d feature vector was extracted during the evaluation phase to calculate gallery and probe distance. Likewise, in Ref. [16] three different gait conditions are examined i.e. normal gait, simulated leg length difference, and simulated leg weight difference. Kinematic temporal changes during gait were captured using an online motion recording system. The resulting hip-knee joint angle diagrams from eight subjects under each gait condition were pre-processed and fed into neural networks. Researchers in Ref. [17] proposed a model-based approach known as Pose Gait that used a deep neural network for the estimation of the 3D human pose of 14 joints and their temporal and spatial features while using 3D human pose images as the input, making it invariant to different views and external variations. However, it achieved only a recognition rate of 49 % for carrying conditions. A model-based automatic cross-view gait recognition system was introduced by Ref. [18], which used shapes and pose sequences extracted directly from the human model. For multi-view constraints, a phased synchronization step was introduced in training to handle inconsistent poses in an up-sampled phase domain. A 3D model-based framework was introduced in Ref. [19], having two branches for extracting features from silhouettes and using the SMPL model to learn 3D shapes and viewpoints. Two new modeling approaches based on coupled oscillators and human locomotion biomechanics were introduced in Ref. [20]. These models estimated thigh and leg inclination from imaging data, which enabled data reduction. Both techniques produced a phase-weighted Fourier representation of the gait characteristic. A model-based method that involves creating initial and posterior models based on anatomical proportions and articulated body parts was proposed by Ref. [21]. Active contour models, the Hough transform, and Fourier analysis are employed to capture motion patterns. The model-based techniques are computationally very complex and require greater computational costs regarding memory and modeling. They also work on 3D data, which is more expensive when compared with traditional methods. Moreover, in Ref. [22] a hybrid gait recognition approach is proposed combining Skeleton Gait Energy Images (SGEI) with Gait Energy Images (GEI). SGEI provides robustness to clothing changes, enhancing GEI's performance. However, limitations include the sensitivity to viewing angles and the need for additional efforts in addressing view-changing challenges.

2.2. Hand crafted appearance-based methods

The problem of known and unknown covariates was tackled in Ref. [14] using a simple CNN model for known covariates and employing discriminative feature learning-based classification using Haralick texture features, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG) as features and SVM, Random Forest, and Multi-layer Perceptron as classifiers for unknown covariates. Although the findings were good, but in this method, the user had to pick the covariate areas each time. A Gaussian Process (GP) classification architecture was proposed to approximate the viewpoint of each probe gait sequence. At various view angles, gait sequences were captured, and then to measure their similarity, Canonical Correlation Analysis (CCA) was used to design the correlation of gait sequences from multiple viewpoints. In contrast, correlation strength was used as a similarity measure [23]. A strategy to address model overfitting due to less training data was proposed by treating it as a bipartite ranking problem and getting inputs from multiple datasets, making it robust for covariate conditions [24]. Researchers introduced a discriminant locally linear embedding

(DLLE) architecture to extract gait features while keeping the native manifold structure [25]. A simple approach was proposed in Ref. [26], which directly extracted features from GEIs by utilizing two subspace learning techniques: discriminant analysis with tensor representation (DATER) and coupled subspace analysis (CSA). Two temporal fusion methods, i.e., image-level fusion and feature-level fusion, were explored to obtain feature representations from a sequence [27]. However, they only achieved 73 % accuracy, and covariates were not handled. Image-level fusion combines information from multiple frames at the image level, while feature-level fusion combines extracted features. In Ref. [28], GEI was divided into eight different sections while assigning different weights to covariate and normal regions. A similar strategy was used in Ref. [29], where human body representation was divided into equal parts, and based on the feature similarity in gait parts, weights were assigned to each part. Likewise, in Ref. [30], a novel method is introduced for gait recognition by dividing the body into segments and identifying effective parts using EnDFT features. It reduced data dimensionality by discarding less effective parts and applying PCA to each effective part. The approach combined individual part distances for classification, outperforming whole-based methods in handling challenging factors like clothing achieving an accuracy of 77 %.

2.3. Automated appearance-based methods

In [31], a novel model is proposed referred to as GEINet consisting of four layers, of which two are convolutional and two are fully connected layers, as it learns gait representation features directly from GEIs while using Softmax loss as its optimizing function. A similar approach was used in Ref. [32], which used multiple pooling and convolution layers to extract gait templates from unordered silhouettes while using batch-all triplet loss for optimization. A deep neural network was introduced in Ref. [33]. It used a framework (TS-CNN) consisting of a two-stream CNN architecture that matches mid-level features by learning similarities among GEI pairs, which it takes as input for recognizing gait at the top layer while having six layers. However, it only worked on normal walk sequences.

Similarly, researchers introduced a part-based network known as Gait Part, which focuses on different body parts' micromotion and their representation [34]. Transfer learning uses knowledge from already-learned networks; for instance, in Ref. [35], a framework is

Table 1
Literature review of latest state-of-the-art methods.

Author	Dataset	Method	Performance	Covariates Handling	Drawbacks
Zhang et al. (2019) [27]	CASIA-B	The combined CNN approach was used to obtain feature representations from sequences through image and feature level fusions.	73 % Accuracy	No	Covariate Conditions are not handled
Liao et al. (2020) [17]	CASIA-B and CASIA-E	PoseGait is proposed in which shapes and pose sequences are extracted directly from the human model.	93.4 % Accuracy for Normal walk and 80.7 % under covariate conditions.	Yes	Computationally expensive being the model-based method
Li et al. (2021) [18]	CASIA-B and OU-MVLP	End-to-end model is proposed for cross-view gait identification method using posture sequences and shape features derived from a human model.	The average recognition rate is 49.72 % for carrying conditions and 96.62 % for normal conditions.	No	Did not handle covariates and was computationally expensive.
Chao et al. (2021) [32]	CASIA-B and OU-MVLP	GaitSet model is proposed on basis of CNN to extract gait templates from a set of unordered silhouettes.	85.9 % accuracy on OU-MVLP dataset and 93.45 % average accuracy for normal walk and 70.3 % for coat wearing sequences on CASIA-B datasets	Yes	Speed conditions were not tested.
Saleem F. et al. (2021) [35]	CASIA-B	NasNet Mobile and the Inception-Resnet-V2 model are used for training, as well as a three-step improved whale optimization approach to choose the best features.	Average Accuracy of 89.0 %	No	Did not handled covariate conditions
Hawas et al. (2019) [43]	CASIA-B	The optical flow of GEI is used to extract static and moving features, which are then input into CNN for classification.	Average Accuracy of 95 %	No	Only identified normal gait sequences, not covariate conditions.
Yao et al. (2021) [22]	CASIA-B	Proposed a hybrid gait recognition technique combining Skeleton Gait Energy Images (SGEI) and Gait Energy Images (GEI).	Average Accuracy of 38 % for Covariate Conditions	Yes	When considering coat wearing conditions and different camera viewpoints, it performs poorly.
Su et al. (2020) [40]	CASIA-B and OU-MVLP	Introduced a novel Center-ranked loss for gait recognition, integrating information from positive and negative samples.	Average Accuracy of 93 % for normal and 65 % for covariate conditions	Yes	Complexity increases with batch size producing scalability challenges.
Bukhari et al. (2021) [14]	CASIA-B and OUR-ISIR	For known covariates, they utilized a basic CNN model, whereas for unknown covariates, they features are extracted using hand-crafted methods followed by using SVM, Random Forest, and MLP classifiers.	Average Accuracy of 96.75 % for normal and 90.32 % for covariate conditions	Yes	Manually selected body parts as static and dynamic regions for covariate recognition

proposed that uses transfer learning for feature extraction. It used NasNet-Mobile and Inception-Resnet-V2 models for training parameters and a three-step enhanced whale optimization algorithm for best feature selection, but the technique only considered known covariate conditions. An event-based gait recognition method known as EV-Gait was proposed by Ref. [36], which was prone to noise in sensory images and used deep neural networks for gait recognition from streams of events. A deep neural network was proposed by Ref. [29], which learns compact and discriminative representations from silhouettes. A joint CNN method was adopted, which produced excellent training, testing, and storage results for gait analysis [27].

A novel framework for person re-identification is proposed, aiming to overcome covariate changes by fusing appearance and gait features [37]. The framework consists of a two-branch model. Firstly, appearance features are extracted from a video sequence using ResNet50, and average pooling is applied to aggregate these features. Secondly, an improved gait representation is designed to capture the person's motion information while mitigating the impact of external covariates. A convolutional neural network (CNN) was proposed for gait recognition during COVID-19 by Ref. [38], which used a parallel CNN architecture during verification. The proposed approach introduces a gait recognition technique that utilizes a multi-layer CNN [39]. Moreover, in Ref. [40], a novel Center-ranked loss is introduced for gait recognition, integrating complete information from positive and negative samples. By employing this loss function in a simple model, it produced good results on CASIA-B and OU-MVLP datasets, particularly in cross-view scenarios and under different conditions. However, the loss function's complexity increases quadratically with batch size, posing scalability challenges, although mitigated by an alternative optimization solution.

For handling multiple covariates, researchers in Ref. [41] introduced an autoencoder that removes invariant gait features. Variable covariate conditions can also be handled using generative adversarial networks (GAN) in gait recognition. An adversarial network known as GaitGAN was introduced by Ref. [42], which removes covariates by generating different feature maps using two discriminator models. Table 1 shows current work on gait identification under normal and covariate settings using the CASIA dataset.

Gait recognition is challenging due to a lack of subject cooperation, making it susceptible to varying covariates across body parts, such as bags and long coats. According to the aforementioned literature research, typical machine learning approaches do not perform well under strict covariate settings since they are initially trained on one covariate and tested on other covariate conditions [26,27]. Similarly, the model-based techniques are computationally very complex and need greater computational cost in terms of memory and modeling. Some studies additionally work on 3D data, which incurs higher costs as compared to standard approaches [17,18,21]. Regions with covariate conditions pose a challenge for deep learning models during the automated feature learning stage owing to intra-class variability. Deep learning models cannot handle stringent covariate challenges since they can integrate information from both static and dynamic areas [14]. Therefore, the identification and handling of covariate regions remain an open research area for end-to-end gait recognition systems.

3. Methodology

The proposed framework for end-to-end gait recognition depicted in Fig. 2, aims to leverage the extraction of both dynamic and static regions. By generating a GEI, a comprehensive representation of the gait patterns is obtained. To effectively capture the discriminative information from the GEI, the model employs parallel filters. These filters operate concurrently on the image, extracting feature maps from both static and dynamic areas.

3.1. Gait energy image (GEI)

Gait Energy Image (GEI) is one of the primary gait representations used in the majority of extant research. The GEI will be calculated from human silhouette sequences using the method proposed in Ref. [44]. All the image frames should have the same size and alignment; hence, size normalization and horizontal alignment will be applied first. Following this, GEI will be calculated from gait cycle silhouettes using the equation below.

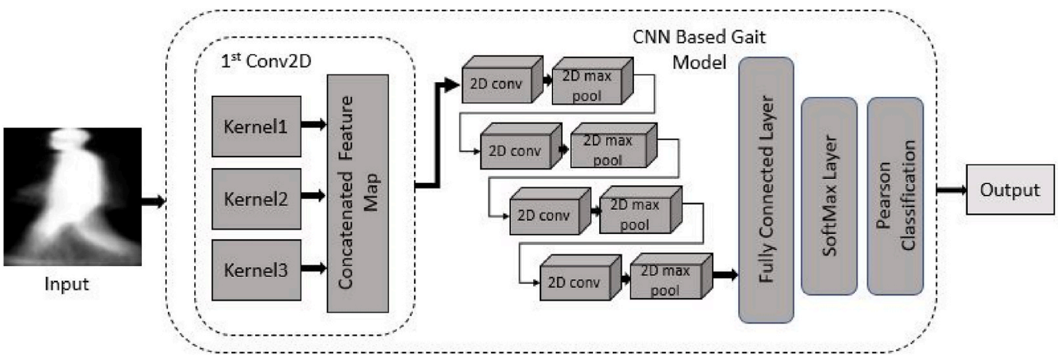

Fig. 2. Pictorial representation of the proposed framework.

$$G(x) = \frac{1}{F}\sum_{f=1}^{F} I(x, y, f), \tag{1}$$

where F represents the total number of frames per gait cycle in Equation (1), both x and y represent the pixel coordinates of the silhouette image I, and f shows the frame number in a gait cycle. Areas that have high intensity represent static areas with not much movement across frames, and they provide feedback regarding the shape and stance of the body. While the regions with low intensity represent dynamic areas that tend to change, like movement while walking [44], the static parts provide vital information when identifying a human. However, they vary when subjected to covariate conditions, such as wearing a bag. In contrast, dynamic regions are not changed much in covariate conditions, making them contain the most crucial information of a GEI as the common covariate conditions like change in clothing or carrying do not affect them, which is normally the typical covariate condition. Therefore, in the presence of variable covariate conditions, dynamic regions are crucial for identifying humans. Static areas, like hairstyle and body structure, also provide handy information for human identification but are open to change in covariate conditions.

3.2. Convolutional neural networks (CNNs)

CNNs have a unique architecture that incorporates convolutional layers, pooling layers, and fully connected layers. The convolutional layers use filters to extract features from the input data, enabling the network to learn hierarchical representations. The pooling layers reduce the spatial dimensions of the feature maps, aiding in capturing important features regardless of their precise locations. Finally, the fully connected layers connect the extracted features to the output layer for classification or regression. In our approach, the grayscale GEI is used as an input with dimensions of 240 × 240 × 1, feeding into the first layer of our CNN. The CNN model consists of 12 layers, with five convolutional layers. The first layer of CNN contains three 3 × 3 filters running in parallel on three horizontal parts of the input. The features from all three kernels are then concatenated to make a feature map, which is further fed to the next convolution and pooling layers. The new feature map will be the same size as the input. The network's weights are updated iteratively using a batch size of 4 and the "Adam" optimization algorithm. Although all kernels are changed, those changes are so minor that when writing in a single decimal, the values remain almost the same. The resulting feature maps for the last four convolutional layers are 16, 32, 64, and 124, respectively, obtained by applying filters to convolve the input image. Throughout the architecture, the LeakyRelu activation function is employed.

$$f(x) = \begin{cases} x, x > 0 \\ 0.05x, \mathit{else} \end{cases} \tag{2}$$

Additionally, the traditional value of $0.01 \times x$ in the LeakyRelu activation function is modified to $0.05 \times x$ in our proposed study, as shown in Equation (2). To reduce the spatial dimensions of the input, max pooling with a 2 × 2 window size is utilized. The fully connected layer takes the resulting information volume and outputs an n-dimensional vector, where n corresponds to the number of classes to be classified. The final output represents the class labels generated by these fully connected layers. It is worth mentioning that the initial convolutional layer filters the input images of size 240 × 240 × 1. Subsequently, the output is split into three regions, each of size (80, 240, 1), covering distinct sections of the image. Convolution operations are then performed on each region with specific kernels, preserving their respective sizes of (80, 240, 1). These regions are merged back into a single tensor, resulting in an output shape of (240, 240, 1) followed by applying convolution and LeakyRelu activation resulting in a (238,238,16) feature maps. Later on, max-pooling with a (2, 2) window size, halving the spatial dimensions and yielding an output shape of (119, 119, 16). In the next step, another convolutional layer with 32 filters using a (3, 3) kernel is applied, followed by max-pooling, reducing the feature map size to (58, 58, 32). Afterward, a subsequent convolutional operation with 64 filters using a (3, 3) kernel followed by (2,2) pooling maintains the feature map size at (28, 28, 64). In the last, the final convolutional layer with 124 filters with a (3, 3) kernel and subsequent max-pooling produces feature maps of size (13, 13, 124). The feature maps are then flattened into a 1D vector comprising 20,956 features, serving as input to a fully connected (FC) layer with 1024 neurons, resulting in an output shape of (1024). Following on, the last layer is the classification layer, which has 124 output units (i.e. equal to the number of classes) and uses softmax activation to assign probabilities to each input and calculate the loss. CNN's learning rate is set to 0.001, the number of epochs is 50, and the kernel size in all convolutions is 3 × 3. To prevent overfitting, dropout layers with a rate of 0.5 are added to the model.

The dynamic regions are more responsible for features that could recognize an individual's gait if compared with static regions, however, the static regions in a GEI have pixel values near the upper bound, which is 255, because when all silhouettes are averaged, the parts not moving all have the same value, hence producing output near 255 or white in the case of static areas. Similarly, the areas of the body that move the most during body movement give gray areas as output when silhouettes are averaged. Legs can be classified as the region that moves the most during a walk, generating the most important dynamic gait features; arms can also be classified as dynamic regions because of movement. However, the torso exhibits static features as it remains idle during movement relative to the viewpoint. Thus, three filters are designed to suppress static features while enhancing features from dynamic areas. The kernels are designed to accommodate both static and dynamic regions. The upper part of GEI consists of the head and shoulders, which do not contribute much to gait analysis being the static areas; thus, a kernel is designed to retain the features without giving them high weight. The middle kernel is designed so that it only encapsulates very limited information due to the majority of the area being static. Similarly, the lower third part contains the most critical information about dynamic regions of the image; hence, a kernel needs to be designed to extract those values from feature maps. This concatenated feature map is then used for further feature learning and extraction. The details of the three kernels are given below.

3.2.1. Kernel 1

The upper third consists of the head and shoulders, which move very little during a walk; hence, kernel one is designed to minimize the feature values. It does not neglect these features completely but dims the region of the input image to minimize the values in this part of the feature map.

$$Kernel\ 1 = \begin{bmatrix} -1 & 0.1 & 0.5 \\ -1 & 1.5 & 0.5 \\ -1 & 0.1 & 0.5 \end{bmatrix} \tag{3}$$

The values for kernel one is shown in Equation (3). It also works as a weak edge detector as the person moves from right to left.

3.2.2. Kernel 2

The middle part consists of the torso and arms, where the torso remains idle, and the arms move to and fro, hence the static and dynamic areas. When covariates are present, i.e., bags and coats, mostly only one side of arm movement is visible in GEI, and the other side gets mixed with the covariate region, making it one. When a person moves in either direction, the part of the arm associated with the dynamic area is the one in front of him, and the one on the backside is associated with the static area.

$$Kernel\ 2 = \begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix} \tag{4}$$

Kernel two is designed in such a way, as shown in Equation (4), to detect edges from left to right. In the dataset, the candidates are moving from right to left; hence, there is a need to design a filter containing only the forward arm part and minimize other features. Thus, an edge detector kernel is used, which will only contain areas covering forward arm motion.

3.2.3. Kernel 3

The lower third part is the most important feature as it contains legs, the most dynamic region in a GEI.

$$Kernel\ 3 = \begin{bmatrix} -0.4 & -0.6 & -0.4 \\ -0.6 & 5.2 & -0.6 \\ -0.4 & -0.6 & -0.4 \end{bmatrix} \tag{5}$$

Hence, the above kernel, as shown in Equation (5), is designed such that it enhances certain gray-level values, which represent dynamic regions, and minimizes the gray levels that are not dominant.

3.3. Concatenating kernels

When the input image is convoluted by these three parallel running kernels, whose output is concatenated to make a new feature map, the next layers learn from that feature map, giving more importance to dynamic areas and less weight to static areas. The new

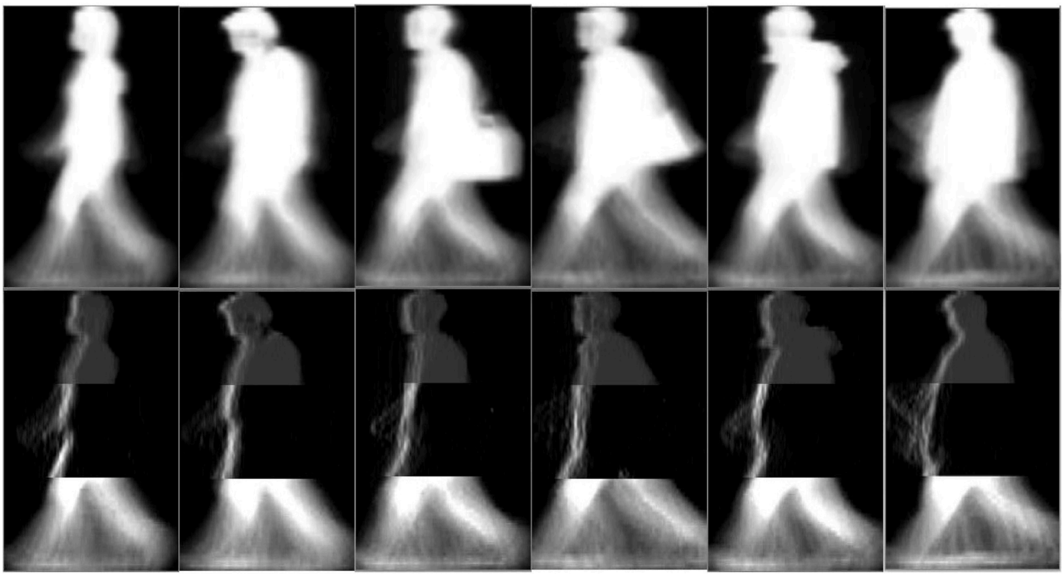

Fig. 3. Example of the original image is provided in the first row and the resultant images produced from the concatenated feature map is provided in second row.

feature map, a kernel-fused map or image, will be of the original image size and produced by combining all three kernel values. This updated feature map will include revised values in all three sections of the original image feature map. Although the kernel values and resultant feature map were updated after every iteration, due to minor changes, the kernel values, when written with one decimal point, almost remained the same. An example of the original and resultant images that should be produced from the feature map after the first layer is given below in Fig. 3.

This convolutional process enhances the understanding of the input image by applying three kernels. The frames are fed to the model sequentially with a customized approach where the input image is divided into three regions, each convolved separately with specific kernels before being merged back into a single tensor. This design likely aims to capture different features from various parts of the image. The resulting outputs from all kernels are then concentrated, creating a new feature map. As the network progresses to subsequent layers, it learns from this feature map, emphasizing dynamic regions that undergo significant changes over time. However, static areas are assigned relatively lower weights, as they tend to remain consistent throughout the gait sequence. The kernels are updated after each iteration in order to minimize the loss. This hierarchical learning approach enables the model to effectively distinguish and prioritize the dynamic components, ultimately leading to improved accuracy and robustness in gait recognition.

4. Experimental setup and results

4.1. Dataset

The proposed framework is evaluated using a widely used benchmark dataset, CASIA. This dataset provides a comprehensive collection of gait sequences captured under various conditions, including normal and covariate conditions. By leveraging this dataset, the framework can be rigorously tested, and its performance is assessed. The CASIA dataset consists of many subject images, with 19,139 images available for analysis. To ensure diversity, the dataset is divided into three subsets: A, B, and C, each serving a specific purpose in evaluating gait recognition algorithms [45]. CASIA subset A comprises normal walk sequences of 20 subjects. This subset allows for assessing the proposed framework's performance under ideal conditions where subjects walk without additional factors or covariates. Subset B of CASIA, conversely, includes walk sequences of 124 subjects with variations in clothing, such as coats and carrying bags. This subset challenges the framework to handle the presence of different attire and items carried by individuals, simulating real-world scenarios. Lastly, CASIA subset C provides normal walk sequences for 153 subjects, covering different walking speeds and even including scenarios where subjects are walking with a bag. This subset introduces variations in walking patterns and behavior, pushing the framework to accurately recognize individuals even in challenging conditions [46]. The rationale for utilizing this dataset is that it comprises a varied variety of covariate conditions, such as subjects carrying bags, coats, and videos at varying walking speeds. These covariate variables provide real-world complexity and challenges that are common in realistic conditions, increasing the dataset's representativeness and the robustness of any models trained on it. In Ref. [14], the CASIA dataset is employed for recognizing normal as well as covariate gait sequences by employing Haralick texture features while in Ref. [46] CASIA A is utilized for recognizing normal walk sequences. Similarly, in Ref. [32] CASIA A for recognizing normal gait, and the CASIA B is utilized for coat-wearing gait sequences. Furthermore, the data was collected in a controlled environment, assuring consistency in lighting and background conditions. This controlled setup helps preserve data quality and allows for successful training and assessment of algorithms.

4.2. Results and discussion

All experiments for this research were conducted on Google Colab using Python, utilizing a two-core CPU running at 2.2 GHz and 13 GB of RAM. The evaluation metric employed in this study was accuracy. The proposed methods were assessed on a standard CASIA gait dataset, considering various cooperative and strict covariate conditions. The results are presented in two sections: gait recognition performance under known covariate conditions and gait recognition performance under strict covariate conditions.

4.2.1. Gait recognition performance under the known covariate conditions

This section focuses on gait recognition results when considering cooperative individuals without covariate conditions. In gait recognition research, it is important to consider known covariate conditions, where the walking conditions in the gallery set are identical to the probe set or a subset. This scenario focuses on training and testing models on similar or identical types of covariates. For instance, a model trained on normal walking sequences can be tested on the same sequences, resulting in known covariates. Another example is to train a model using gait data acquired while wearing a coat and then test it under similar conditions. The GEI images were

Table 2
Performance of traditional CNN and custom kernel CNN on normal conditions.

Gallery	Probe	CNN Accuracy	Custom Kernel CNN Accuracy
Normal	Normal	90.33 %	95.00 %
Normal to Coats	Normal	87.00 %	94.00 %
Normal Speed	Normal Speed	99.00 %	99.67 %
Fast Speed	Normal Speed	93.00 %	97.38 %
Slow Speed	Normal Speed	94.00 %	96.73 %

put directly into a CNN algorithm and assessed using a customized kernel model, the features of which were then concatenated to create a new feature map, yielding very accurate results as shown in Table 2. The feature map and kernel values are modified after every run to reduce loss. Although we employed customized kernels, a basic CNN may provide acceptable outcomes with known covariates. Table 2 compares the accuracy of the standard CNN with a suggested custom kernel CNN. Notably, speed is not taken into account as a strict covariate condition since differences in walking speed have no substantial effect on the human body shape. Fig. 4 depicts the model's accuracy for speed changes, whereas Fig. 5 depicts the model's loss for speed variations. The model was trained on normal and evaluated on normal, fast, and slow walk sequences. Strict covariate conditions are limited to bag and clothing variations. Although the custom kernel-based method achieved higher accuracy, this also supports the notion that GEI is effective for gait recognition under normal circumstances. However, it is important to note that the performance significantly deteriorates in the presence of covariate conditions, as widely reported, making the custom kernel model a more suitable choice in such scenarios.

Table 3 compares the performance of the proposed method with existing approaches to gait recognition, focusing on the ability to recognize individuals without considering covariates such as bags, clothing, etc. By evaluating the proposed deep learning method against the existing work, the study aims to provide insights into its potential advantages and limitations in gait recognition under no covariate conditions. Our experimental findings demonstrate that our proposed method outperforms existing approaches to handling non-covariate conditions. These results lead to an important conclusion: a straightforward implementation of GEI with deep learning can address non-covariate conditions effectively and does not need a custom kernel-based method.

4.2.2. Gait recognition performance under strict covariate conditions

This section discusses the findings obtained under strict covariate conditions, where the gallery and probe sets do not overlap in terms of walking scenarios. To solve the issues created by these covariate scenarios a straightforward approach is taken. It involves eliminating only relevant and important regions of interest (ROIs) from the GEI (Gait Energy Image) or giving priority to dynamic areas. This selective extraction allows us to focus solely on the common parts shared by the GEIs of both the gallery and probe sets. Table 4 presents the outcomes for the CASIA dataset, where training is conducted solely on normal GEIs while testing is carried out separately on bag and coat GEIs. It is evident from Table 4 that traditional CNN struggles to produce satisfactory results when strict covariate conditions are present. Table 4 also includes a comparative analysis between traditional CNN and a custom kernel-based CNN, and this shows that proposed custom-kernel CNN model performs better than traditional CNN. The reason for this improvement is that proposed model intends to eliminate the features causing degradation in performance i.e. features resultant from covariate regions. The proposed model employs three custom kernels to emphasize those features which are discriminative enough to be utilized for gait recognition. The regions including covariate factors such as bags are eliminated and ignored through custom kernels, and they only extract features from non-covariate regions while the traditional CNN-based method extract features including covariate regions which make feature set less discriminative to be utilized for accurate gait recognition.

Figs. 6 and 7 illustrate the model's accuracy and loss when evaluated under strict covariate settings, including bags and coats. When bags are included as covariate conditions, the model's overall training and validation accuracy improves as the epoch's progress.

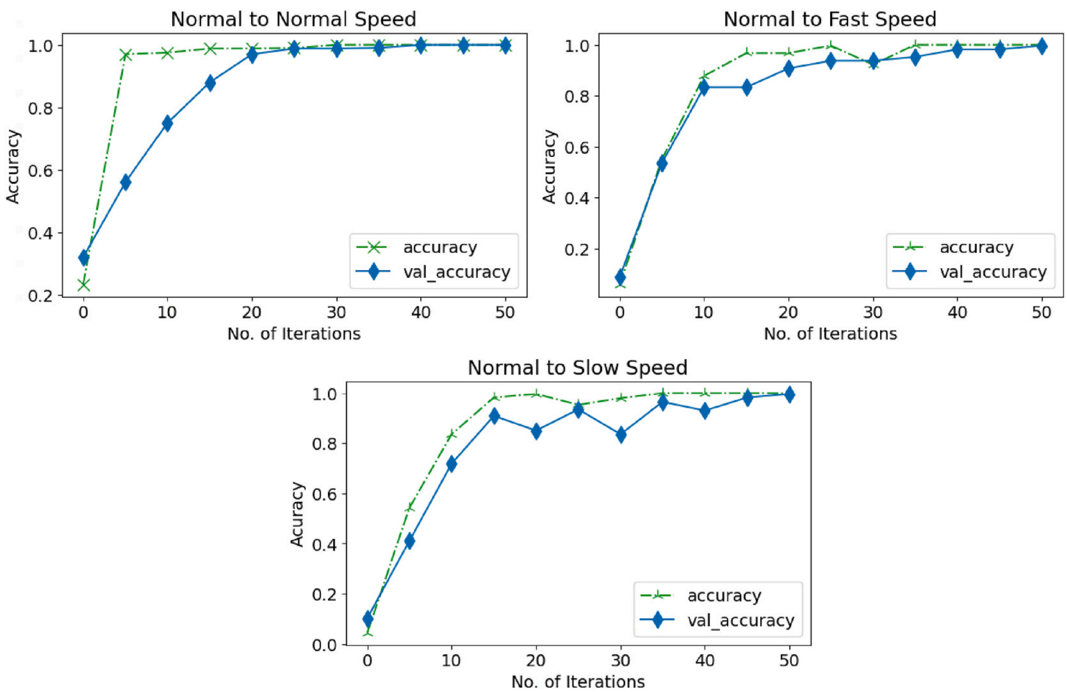

Fig. 4. Model accuracy over speed variations.

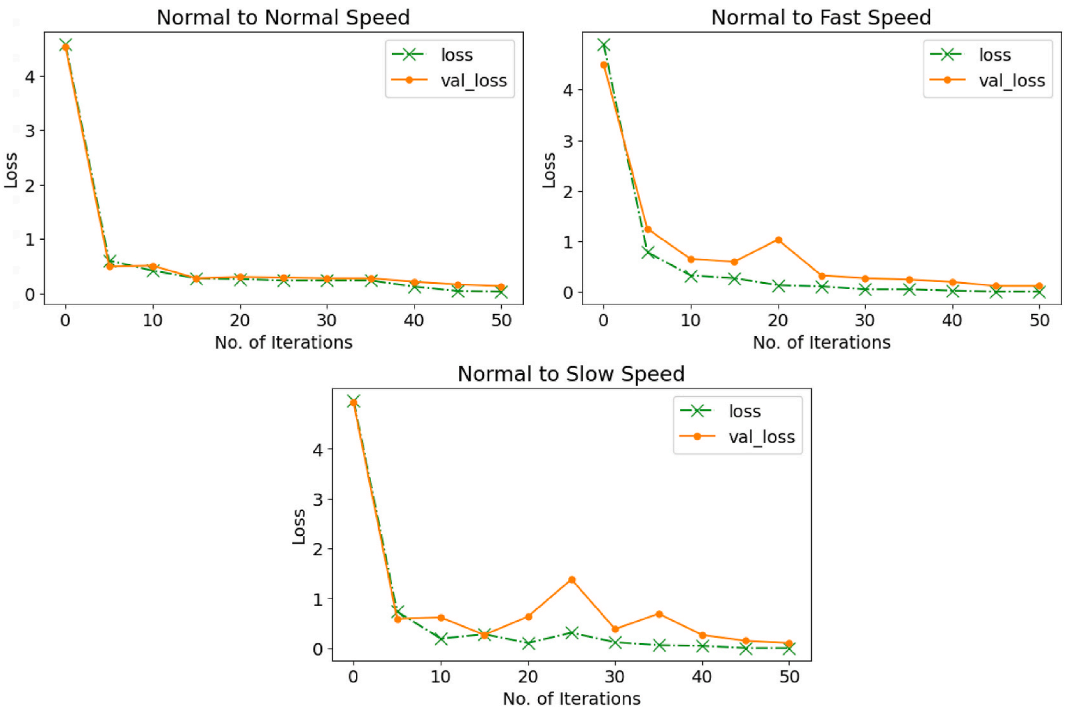

Fig. 5. Model loss over speed variations.

Table 3
Comparison between different techniques in the literature for normal conditions.

Authors	Technique	Covariate	Average Accuracy
Hawas et al. [43]	Deep Learning	No	93.50 %
Liao et al. [17]	Deep Learning	No	63.78 %
Lishani et al. [47]	Deep Learning	No	92 %
Wu et al. [48]	Deep Learning	No	95.96 %
Proposed Method	**Deep Learning**	**Yes**	**96.56 %**

Table 4
Comparison between traditional and custom kernel CNN for covariate conditions.

Gallery	Probe	CNN Accuracy	Custom Kernel CNN Accuracy
Normal	Bags	83 %	90 %
Normal	Coats	52 %	58 %

Similarly, when dealing with coat-wearing scenarios the model shows a rapid decrease in both training and validation loss, indicating effective learning.

To assess the effectiveness of our method, we compared our results with techniques designed for strict covariate conditions. Our findings demonstrate that incorporating bag or coat sequences in the gallery set can yield favorable results. This set was particularly evident in our first experiment, where we used GEI with CNN and achieved impressive outcomes. Moreover, we compared only available data without employing data augmentation techniques. Table 5 shows that our method produced comparable findings with respect to cutting-edge methods, showing good performance across a variety of criteria, particularly for unknown covariate factors i.e. wearing bags. By prioritizing dynamic regions and suppressing static ones, the model efficiently captures discriminative information, resulting in increased accuracy. The experimental findings show higher performance, particularly under severe covariate circumstances such as bags and coats. This demonstrates its ability to solve difficult challenging conditions and compete effectively.

5. Conclusion

Covariate variables, such as carrying and wearing conditions, variances in walking speed, and variations in camera viewpoints, can all have an impact on gait recognition system performance, potentially resulting in lower accuracy. Traditional machine learning

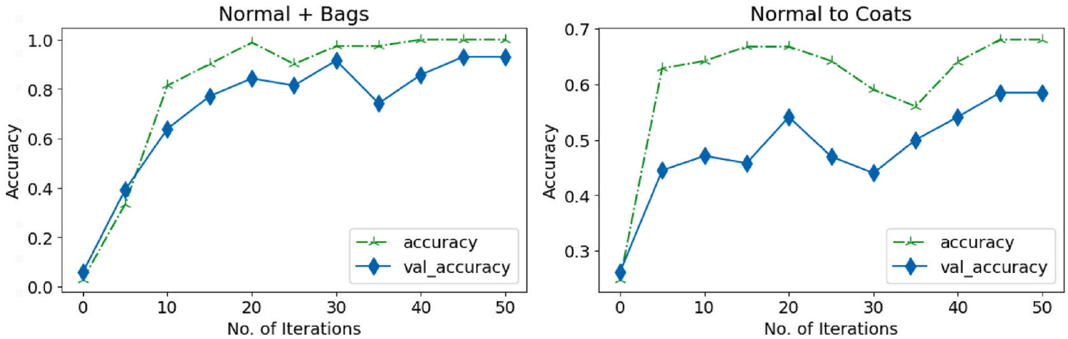

Fig. 6. Model Accuracy over bag (carrying) and coat (wearing) conditions.

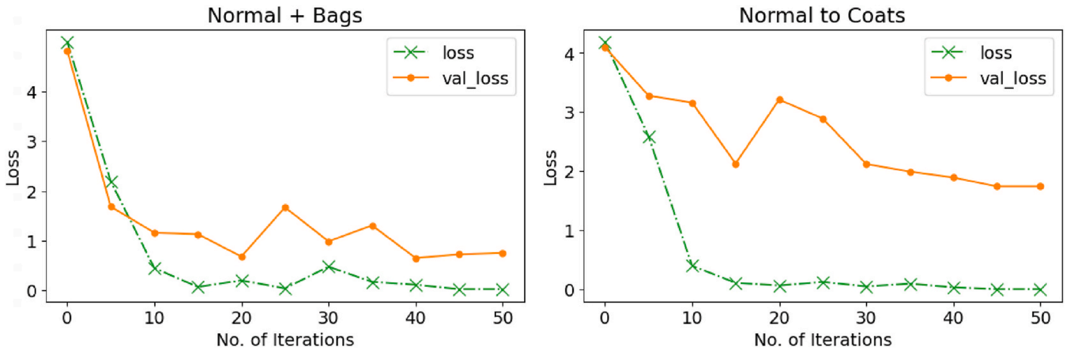

Fig. 7. Model Loss over bag (carrying) and coat (wearing) conditions.

Table 5
Comparison of several approaches on a strict covariate condition.

no.	Method	Normal	Bags	Coats	Avg. Accuracy	Covariates
1	Bashir et al. [28]	100 %	78.30 %	44.40 %	74.20 %	Yes
2	Gupta et al. [49]	NA	86.20 %	61.40 %	73.80 %	Yes
3	Hawas et al. [43]	97.60 %	45.30 %	49.60 %	64.10 %	Yes
4	Yu et al. [41]	95.97 %	65.32 %	42.74 %	68.01 %	Yes
5	Rokanujjaman et al. [30]	97.61 %	83.87 %	51.61 %	77.69 %	Yes
6	Yao et al. [22]	NA	NA	38.00 %	38.00 %	Yes
7	Su et al. [40]	93.20 %	72.80 %	59.10 %	75.03 %	Yes
8	Huang et al. [50]	91.90 %	80.30 %	72.30 %	81.50 %	Yes
9	Bukhari et al. [14]	97.98 %	81.00 %	92.00 %	90.32 %	Yes
	Proposed Method	**96.56 %**	**90.00 %**	**58.00 %**	**81.52 %**	**Yes**

methods perform poorly under stringent covariate situations because they are trained on handcrafted features. As a result, this paper proposes a unique deep learning strategy for detecting subjects with covariate conditions, which extracts feature maps using a CNN by employing customized kernels on distinct image regions followed by concatenating these feature maps to make a fused feature map. More precisely, the proposed method determines static and dynamic areas of interest where static areas contain covariates, and then features will be learned from the dynamic regions unaffected by covariates to recognize the individuals efficiently. Three 3×3 kernels are proposed in the first layer of the CNN to propose regions of interest having covariate conditions from GEIs, as these regions exhibit more information about the gait pattern of the subject. The experiment was conducted on the CASIA dataset, and the findings show that identifying individuals wearing bags had 90 % accuracy, while those wearing coats had 58 % accuracy, indicating an improvement over earlier deep learning approach. The suggested technique addressed covariates with bags with 90 % accuracy but only 58 % accuracy for covariates with coats; however, the loss can be lowered in future work by employing transfer learning approaches. In future research, dedicated efforts can be made to design a standard architecture that seamlessly integrates transfer learning and reinforcement learning, enabling robust gait recognition that handles all covariates.

CRediT authorship contribution statement

Babar Ali: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Maryam Bukhari:** Methodology, Investigation, Formal analysis, Data curation. **Muazzam Maqsood:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Formal analysis. **Jihoon Moon:** Visualization, Validation, Formal analysis. **Eenjun Hwang:** Project administration, Funding acquisition. **Seungmin Rho:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2024(Project Name: Development of distribution and management platform technology and human resource development for blockchain-based SW copyright protection, Project Number:RS-2023-00228867, Contribution Rate: 100 %)

References

[1] V. Rai, K. Mehta, J. Jatin, D. Tiwari, R. Chaurasia, Automated biometric personal identification-techniques and applications, in: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, Madurai, India, May 2020, pp. 1023–1030, https://doi.org/10.1109/ICICCS48265.2020.9120896.
[2] M, Whittle & Levine, David & Richards, Jim & G, Barton, "Applications of gait analysis.," in Whittle's Gait Analysis., fifth ed., pp. 113–123.
[3] C. Shen, S. Yu, J. Wang, G.Q. Huang, L. Wang, A Comprehensive Survey on Deep Gait Recognition: Algorithms, Datasets and Challenges, 2022, https://doi.org/10.48550/ARXIV.2206.13732.
[4] M. Kumar, N. Singh, R. Kumar, S. Goel, K. Kumar, Gait recognition based on vision systems: a systematic survey, J. Vis. Commun. Image Represent. 75 (Feb. 2021) 103052, https://doi.org/10.1016/j.jvcir.2021.103052.
[5] V.B. Semwal, M. Raj, G.C. Nandi, Biometric gait identification based on a multilayer perceptron, Robot. Auton. Syst. 65 (Mar. 2015) 65–75, https://doi.org/10.1016/j.robot.2014.11.010.
[6] J.C. Gonzalez-Islas, O.A. Dominguez-Ramirez, H. Castillejos-Fernandez, F.A. Castro-Espinoza, Human gait analysis based on automatic recognition: a review, Pädi Bol. Científico Cienc. Básicas E Ing. ICBI 10 (Especial3) (Aug. 2022) 13–21, https://doi.org/10.29057/icbi.v10iEspecial3.8927.
[7] Xiaxi Huang, N.V. Boulgouris, Model-based human gait recognition using fusion of features, in: *2009 IEEE International Conference On Acoustics, Speech And Signal Processing*, Taipei, Taiwan, IEEE, Apr. 2009, pp. 1469–1472, https://doi.org/10.1109/ICASSP.2009.4959872.
[8] D. Sethi, S. Bharti, C. Prakash, A comprehensive survey on gait analysis: history, parameters, approaches, pose estimation, and future work, Artif. Intell. Med. 129 (Jul. 2022) 102314, https://doi.org/10.1016/j.artmed.2022.102314.
[9] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2) (Feb. 2006) 316–322, https://doi.org/10.1109/TPAMI.2006.38.
[10] A. Parashar, R.S. Shekhawat, W. Ding, I. Rida, Intra-class variations with deep learning-based gait analysis: a comprehensive survey of covariates and methods, Neurocomputing 505 (Sep. 2022) 315–338, https://doi.org/10.1016/j.neucom.2022.07.002.
[11] H. Iwama, M. Okumura, Y. Makihara, Y. Yagi, The OU-isir gait database comprising the large population dataset and performance evaluation of gait recognition, IEEE Trans. Inf. Forensics Secur. 7 (5) (Oct. 2012) 1511–1521, https://doi.org/10.1109/TIFS.2012.2204253.
[12] K.T. Thomas, K.P. Pushpalatha, A comparative study of the performance of gait recognition using gait energy image and shannon's entropy image with CNN, in: S. Shukla, A. Unal, J.V. Kureethara, D.K. Mishra, D.S. Han (Eds.), Data Science and Security, Lecture Notes in Networks and Systems, vol.290, Springer Singapore, Singapore, 2021, pp. 191–202, https://doi.org/10.1007/978-981-16-4486-3_21, 290.
[13] A. Sepas-Moghaddam, A. Etemad, Deep gait recognition: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 45 (1) (Jan. 2023) 264–284, https://doi.org/10.1109/TPAMI.2022.3151865.
[14] M. Bukhari, et al., An efficient gait recognition method for known and unknown covariate conditions, IEEE Access 9 (2021) 6465–6477, https://doi.org/10.1109/ACCESS.2020.3047266.
[15] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hormann, G. Rigoll, Gaitgraph: graph convolutional network for skeleton-based gait recognition, in: *2021 IEEE International Conference On Image Processing (ICIP)*, Anchorage, AK, USA: IEEE, Sep. 2021, pp. 2314–2318, https://doi.org/10.1109/ICIP42928.2021.9506717.
[16] J.G. Barton, A. Lees, An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams, Gait Posture 5 (1) (Feb. 1997) 28–33, https://doi.org/10.1016/S0966-6362(96)01070-3.
[17] R. Liao, S. Yu, W. An, Y. Huang, A model-based gait recognition method with body pose and human prior knowledge, Pattern Recognit 98 (Feb. 2020) 107069, https://doi.org/10.1016/j.patcog.2019.107069.
[18] X. Li, Y. Makihara, C. Xu, Y. Yagi, End-to-end model-based gait recognition using synchronized multi-view pose constraint, in: *2021 IEEE/CVF International Conference On Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, IEEE, Oct. 2021, pp. 4089–4098, https://doi.org/10.1109/ICCVW54120.2021.00456.
[19] J. Zheng, X. Liu, W. Liu, L. He, C. Yan, T. Mei, Gait recognition in the wild with dense 3D representations and A benchmark, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New Orleans, LA, USA, Jun. 2022, pp. 20196–20205, https://doi.org/10.1109/CVPR52688.2022.01959.
[20] C. Yam, M.S. Nixon, J.N. Carter, Automated person recognition by walking and running via model-based approaches, Pattern Recognit 37 (5) (May 2004) 1057–1072, https://doi.org/10.1016/j.patcog.2003.09.012.
[21] F. Tafazzoli, R. Safabakhsh, Model-based human gait recognition using leg and arm movements, Eng. Appl. Artif. Intell. 23 (8) (Dec. 2010) 1237–1246, https://doi.org/10.1016/j.engappai.2010.07.004.
[22] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, Z. Tang, W. Yang, Robust gait recognition using hybrid descriptors based on Skeleton Gait Energy Image, Pattern Recognit. Lett. 150 (Oct. 2021) 289–296, https://doi.org/10.1016/j.patrec.2019.05.012.
[23] K. Bashir, T. Xiang, S. Gong, Cross view gait recognition using correlation strength, in: Procedings of the British Machine Vision Conference 2010, British Machine Vision Association, Aberystwyth, 2010, pp. 109.1–109.11, https://doi.org/10.5244/C.24.109.
[24] R. Martín-Félez, T. Xiang, Uncooperative gait recognition by learning to rank, Pattern Recognit 47 (12) (Dec. 2014) 3793–3806, https://doi.org/10.1016/j.patcog.2014.06.010.
[25] Xuelong Li, S. Lin, Shuicheng Yan, Dong Xu, Discriminant locally linear embedding with high-order tensor data, IEEE Trans. Syst. Man Cybern. Part B Cybern. 38 (2) (Apr. 2008) 342–352, https://doi.org/10.1109/TSMCB.2007.911536.
[26] Dong Xu, Shuicheng Yan, Dacheng Tao, Lei Zhang, Xuelong Li, Hong-Jiang Zhang, Human gait recognition with matrix representation, IEEE Trans. Circuits Syst. Video Technol 16 (7) (Jul. 2006) 896–903, https://doi.org/10.1109/TCSVT.2006.877418.

[27] Y. Zhang, Y. Huang, L. Wang, S. Yu, A comprehensive study on gait biometrics using a joint CNN-based method, Pattern Recognit 93 (Sep. 2019) 228–236, https://doi.org/10.1016/j.patcog.2019.04.023.

[28] K. Bashir, Tao Xiang, Shaogang Gong, Gait recognition using gait entropy image, in: 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), IET, London, UK, 2009, https://doi.org/10.1049/ic.2009.0230. P2–P2.

[29] S. Hou, C. Cao, X. Liu, Y. Huang, Gait lateral network: learning discriminative and compact representations for gait recognition, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision – ECCV 2020, Lecture Notes in Computer Science, vol.12354, Springer International Publishing, Cham, 2020, pp. 382–398, https://doi.org/10.1007/978-3-030-58545-7_22, 12354.

[30] Md Rokanujjaman, Md S. Islam, Md A. Hossain, Md R. Islam, Y. Makihara, Y. Yagi, Effective part-based gait identification using frequency-domain gait entropy features, Multimed. Tools Appl. 74 (9) (May 2015) 3099–3120, https://doi.org/10.1007/s11042-013-1770-8.

[31] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi, GEINet: view-invariant gait recognition using a convolutional neural network, in: *2016 International Conference On Biometrics (ICB)*, Halmstad, Sweden, IEEE, Jun. 2016, pp. 1–8, https://doi.org/10.1109/ICB.2016.7550060.

[32] H. Chao, K. Wang, Y. He, J. Zhang, J. Feng, GaitSet: cross-view gait recognition through utilizing gait as a deep set, IEEE Trans. Pattern Anal. Mach. Intell. (2021), https://doi.org/10.1109/TPAMI.2021.3057879, 1–1.

[33] Z. Wu, Y. Huang, L. Wang, X. Wang, T. Tan, A comprehensive study on cross-view gait based human identification with deep CNNs, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2) (Feb. 2017) 209–226, https://doi.org/10.1109/TPAMI.2016.2545669.

[34] C. Fan, et al., GaitPart: temporal part-based model for gait recognition, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, Jun. 2020, pp. 14213–14221, https://doi.org/10.1109/CVPR42600.2020.01423.

[35] F. Saleem, et al., Human gait recognition: a single stream optimal deep learning features fusion, Sensors 21 (22) (Nov. 2021) 7584, https://doi.org/10.3390/s21227584.

[36] Y. Wang, et al., EV-gait: event-based robust gait recognition using dynamic vision sensors, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, Jun. 2019, pp. 6351–6360, https://doi.org/10.1109/CVPR.2019.00652.

[37] X. Lu, X. Li, W. Sheng, S.S. Ge, Long-term person Re-identification based on appearance and gait feature fusion under covariate changes, Processes 10 (4) (Apr. 2022) 770, https://doi.org/10.3390/pr10040770.

[38] M.S.S. Sayeed, P.P. Min, M.A. Bari, Deep learning based gait recognition using convolutional neural network in the COVID-19 pandemic, Emerg. Sci. J 6 (5) (Aug. 2022) 1086–1099, https://doi.org/10.28991/ESJ-2022-06-05-012.

[39] J.H. Yuan, Research on gait feature fusion method of OpenPose and GEI based on convolutional network, in: 2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE), IEEE, Guangzhou, China, Feb. 2023, pp. 175–178, https://doi.org/10.1109/NNICE58320.2023.10105757.

[40] J. Su, Y. Zhao, X. Li, Deep metric learning based on center-ranked loss for gait recognition, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Barcelona, Spain, May 2020, pp. 4077–4081, https://doi.org/10.1109/ICASSP40776.2020.9054645.

[41] S. Yu, H. Chen, Q. Wang, L. Shen, Y. Huang, Invariant feature extraction for gait recognition using only one uniform model, Neurocomputing 239 (May 2017) 81–93, https://doi.org/10.1016/j.neucom.2017.02.006.

[42] S. Yu, H. Chen, E.B.G. Reyes, N. Poh, GaitGAN: invariant gait feature extraction using generative adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Honolulu, HI, USA, Jul. 2017, pp. 532–539, https://doi.org/10.1109/CVPRW.2017.80.

[43] A.R. Hawas, H.A. El-Khobby, M. Abd-Elnaby, F.E. Abd El-Samie, Gait identification by convolutional neural networks and optical flow, Multimed. Tools Appl. 78 (18) (Sep. 2019) 25873–25888, https://doi.org/10.1007/s11042-019-7638-9.

[44] K. Bashir, T. Xiang, S. Gong, Gait recognition without subject cooperation, Pattern Recognit. Lett. 31 (13) (Oct. 2010) 2052–2060, https://doi.org/10.1016/j.patrec.2010.05.027.

[45] Z. Shuai. CASIA, 2008 [Online]. Available: http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp.

[46] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, Robust view transformation model for gait recognition, in: 2011 18th IEEE International Conference on Image Processing, IEEE, Brussels, Belgium, Sep. 2011, pp. 2073–2076, https://doi.org/10.1109/ICIP.2011.6115889.

[47] A.O. Lishani, L. Boubchir, E. Khalifa, A. Bouridane, Human gait recognition using GEI-based local multi-scale feature descriptors, Multimed. Tools Appl. 78 (5) (Mar. 2019) 5715–5730, https://doi.org/10.1007/s11042-018-5752-8.

[48] Wu, X., T. Yang, and Z. Xia, “"Gait recognition based on densenet transfer learning.,” Int. J. Sci. Environ., pp. 1–14.

[49] S.K. Gupta, G.M. Sultaniya, P. Chattopadhyay, An efficient descriptor for gait recognition using spatio-temporal cues, in: J.K. Mandal, D. Bhattacharya (Eds.), Emerging Technology in Modelling and Graphics, Advances in Intelligent Systems and Computing, vol.937, Springer Singapore, Singapore, 2020, pp. 85–97, https://doi.org/10.1007/978-981-13-7403-6_10, 937.

[50] G. Huang, Z. Lu, C.-M. Pun, L. Cheng, Flexible gait recognition based on flow regulation of local features between key frames, IEEE Access 8 (2020) 75381–75392, https://doi.org/10.1109/ACCESS.2020.2986554.